Test Methods for Human-System Interaction

Jean Scholtz

Three different types of Human-system interaction requirements

- Check lists
  - Is the following information contained on the dashboard?
- User testing
  - Acceptable usability
  - Operator to robot ratio
- Experimental constraints
  - Must be able to operate the robot while walking
  - Must be able to operate the robot wearing protective clothing
- Combination of checklists, user testing
  - Is initial training included?, is acceptable usability achieved after initial training?

What would a user test consist of?

- Acceptable usability
  - Would need a representative user
  - Would pick a terrain representative of the scenario
  - Will select appropriate conditions for the test
    - Daylight, nighttime
    - Wearing protective clothing
    - While moving
  - Need to define “primitive tasks”
    - “navigate to a specified location”
    - “operate the cameras”
      - Pan/tilt, zoom, switch between cameras
    - “report location of victim”
  - For each task, we will define metrics
    - Accuracy, time, workload are examples

What would an empirical evaluation look like? Cont.

- We would need 5-8 users for each test
  - Per robot
- We need to determine
  - The primitive tasks (appropriate granularity)
  - Determine the metrics for each
    - And determine what the acceptable level is for each metric
- We should also consider
  - Evaluating acceptable usability for other roles
    - Structural engineer, hazmat engineer
    - How accurately and in what time frame can they make a decision based on the information given to them?

How does this get reported?

- Common Industry Format (CIF)
  - ANSI NCITS 354-2001
    - Just approved as an ISO standard
  - Specifies that one reports
    - Demographics/ number of users
    - Testing conditions
    - Hardware & software descriptions
      - We would modify this to include a description of the OCU
    - Tasks that users were asked to do
      - We will specify what these are
    - How the data was collected and analyzed
    - Results
      - Efficiency, effectiveness, user satisfaction
      - We would replace with the appropriate metrics for each task

What needs to be done?

- We need to work with small groups of first responders to identify primitive tasks
  - Example: navigate from point A to point B
- And we need to identify the metrics that make sense for each primitive tasks
- We need to determine how to score the results for the set of primitive tasks
  - Are all task weighted equally? Are all metrics equal?
- We need to define the skills that we expect operators to have
  - So we recruit the correct set of users for testing
- We need to define the physical setup for the experimental testing
  - How much realism is needed ? Laboratory versus field testing
- We need to then conduct a number of these tests using various platforms and various operators to validate the experimental methods and metrics